Energy 285 (2023) 129513

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A new data mining strategy for performance evaluation of a shared energy recovery system integrated with data centres and district heating networks

Han Du [a], Xinlei Zhou [a], Natasa Nord [b], Yale Carden [c], Zhenjun Ma [a,*]

[a] Sustainable Buildings Research Centre, University of Wollongong, NSW, 2522, Australia
[b] Department of Energy and Process Engineering, Norwegian University of Science and Technology, Norway
[c] Geoexchange Australia PTY Ltd., North Sydney, NSW, 2060, Australia

ARTICLE INFO

Handling Editor: X Zhao

Keywords:
Data centre
Waste heat utilization
District heating
Multidimensional and temporal clustering
Quantitative analysis

ABSTRACT

This study presents a new data mining strategy to discover the performance and operational patterns of a shared energy recovery (SER) system with a data centre and a district heating network. Multidimensional clustering incorporated with a composite performance metric was first used to evaluate the typical performance of the system and reveal the interactions among different performance indicators. Decision tree analysis was then used to identify distinct system performances under different external conditions. Temporal clustering analysis was lastly used to identify the impact of recovered waste heat on the variations in heat supply from the district heating substation. The strategy was evaluated through a case study SER system at a university campus located in Norway. It was found that the most frequent performance accounted for 34 % of the total operational period with the instantaneous waste heat recovery rate of 572.9 kW, the temperature of waste heat of 57.2 °C, and the coefficient of performance of the heat pumps of 2.0. The outdoor air temperature and supply water temperature from the main district heating substation to the campus buildings showed a significant impact on the SER system performance. Moreover, the results showed that the SER system can help reduce the energy use of the district heating networks while increasing the fluctuations of heat supply from the main district heating substation.

1. Introduction

According to the 2020 United Nations Global Status Report, the operation and construction of buildings generated 38 % of total CO_2 emissions and consumed 35 % of global end-energy use [1]. To address these concerns, the United Nations established the Sustainable Development Goals (SDGs) to guide global development efforts with a focus on achieving sustainable cities and communities that prioritize energy conservation and renewable energy usage, and reduce greenhouse gas emissions [2]. Consequently, assessing the performance of energy systems used in buildings becomes essential to improve their energy efficiency, reduce energy use, and minimize greenhouse gas emissions to achieve sustainable development [3].

There is growing interest in shared energy recovery (SER) systems as a technology solution to assist with achieving the SDGs. As shown in Fig. 1, such a system can capture waste thermal energy from one or more sources, such as data centres, ice rinks, or industrial processes, and redistribute the captured waste energy to nearby users for heating, cooling, electricity generation, and/or other purposes. The SER system is a promising way to improve energy efficiency and reduce greenhouse gas emissions, as it can reuse the waste thermal energy that would otherwise be wasted.

The potential benefits of the SER system have been demonstrated in several recent studies [4]. Abdalla et al. [5], for example, connected a small cluster of buildings, including an ice hockey arena, a library with an IT server, and recreation centres with a swimming pool and recreation activities, with a low-temperature network for space heating and cooling, which can effectively allow thermal energy to be shared among the facilities with minimum thermal energy losses. The results showed that this SER system could cover 48 % of the total heating demand and could further cover an additional 12 % of thermal energy when short-term thermal storage was used, leading to a 74 % reduction in total greenhouse gas emissions. Wirtz et al. [6] presented an optimal design approach based on Linear Programming to design bidirectional low-temperature networks that incorporate various cooling and heating sources, such as chillers, pumps, heat exchangers, boilers, hot and cold thermal energy storage systems, and batteries, as well as renewable energy resources such as solar photovoltaics. By optimizing the selection and size of system components, it was found that the optimized system

* Corresponding author.
E-mail address: zhenjun@uow.edu.au (Z. Ma).

https://doi.org/10.1016/j.energy.2023.129513
Received 24 June 2023; Received in revised form 26 September 2023; Accepted 28 October 2023
Available online 29 October 2023

Nomenclature

Greek letters

α	Average value
β	Ratio of the average value to the maximum value
μ	A set of cluster centres

Abbreviations

CART	Classification and regression trees
COP	Coefficient of performance
DH	District heating
HE	Heat exchanger
HP	Heat pump
HVAC	Heating, Ventilation and Air Conditioning
ORC	Organic Rankine cycle
PAR	Peak-to-average ratio
SER	Shared energy recovery
SDG	Sustainable Development Goals

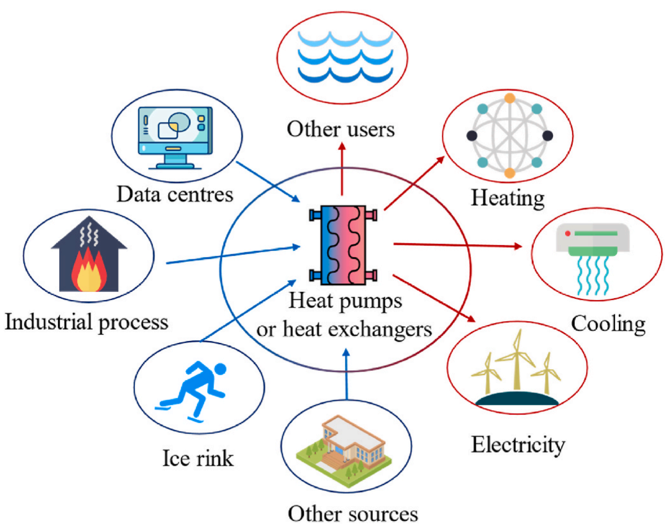


Fig. 1. Schematic of an SER system.

could reduce annual costs and carbon emissions by 42 % and 56 % respectively, as compared to stand-alone Heating, Ventilation and Air Conditioning (HVAC) systems. In Ref. [7], heat pumps were used to capture waste heat from the cooling system of a data centre that requires year-round cooling, and the recovered waste heat was shared with a multi-unit residential building located in a cold climate. The results showed that energy sharing reduced the heating requirements of the building by 55 % and greenhouse gas emissions by 53 %. Meanwhile, it reduced the cooling requirements and greenhouse gas emissions of the data centre by 50 % and 51 %, respectively.

Data centres have the potential to generate a significant amount of waste heat and can be used as a valuable heat resource in an SER system [8]. One common application is to feed the heat recovered into district heating (DH) networks [9]. For example, reusing waste heat from a 3.5 MW data centre as part of DH networks in London could lead to over 4000 tonnes of CO_2 reduction and nearly £1 million in annual cost savings [10]. The impact of reusing waste heat in DH networks was quantified through simulations by Wahlroos et al. [11]. It was found that 0.6–7.3 % of operational costs could be saved when the amount of waste heat ranging from 20 to 60 MW was provided by the data centres for the DH network. However, a potential challenge is a mismatch between the thermal energy supplied by the data centres and the heating requirements of the DH consumers on a daily and seasonal basis [9,12]. To address this issue, Li et al. [13] introduced two types of thermal energy storage, including a water tank for short-term storage and a borehole thermal energy storage system for long-term storage. For short-term storage, the water tank can shave the peak load by 31 % and reduce the annual energy cost by 5 %. For long-term storage, the borehole system can increase the rate of waste heat utilization from 77 % to 96 % and reduce CO_2 emissions by 8 % annually. These findings highlighted the potential of SER systems integrated with data centres to increase energy efficiency and flexibility, reduce carbon emissions, and enhance sustainability in various applications.

In addition to providing waste heat to DH systems, waste heat from data centres can also be used to power thermally driven air conditioning systems or generate electricity based on the Organic Rankine cycle (ORC) [14]. For example, the performance of a silica gel-water adsorption chiller driven by thermal energy shared from a data centre was studied by Pan et al. [15]. The experimental results showed that the coefficient of performance (COP) of this system was in the range of 0.283–0.477, which was relatively low compared to that of traditional adsorption chillers. Araya et al. [16] studied a lab-scale ORC driven by low-grade waste heat from a server rack (40–85 °C) through experimental and theoretical analysis and it was shown that thermal efficiencies varied from 1.9 % at 43 °C to 4.6 % at 81 °C. These studies indicated the potential of using waste heat from data centres to power thermally driven systems and generate electricity, thereby contributing to a more sustainable energy mix.

Based on the above analysis, it can be concluded that SER systems showed great promise for achieving building energy efficiency and sustainability. However, the majority of the existing studies used simulations, which often cannot consider practical dynamics and constraints in the real-time operation of such integrated systems. To the best of our knowledge, very limited studies used field-measured data to analyze the performance of SER systems. Data mining has shown great potential to discover useful information from large datasets [17–21] and has not previously been used to explore energy performance characteristics and energy-saving opportunities of SER systems integrated with data centres and DH systems.

To this end, this paper presents a new data mining strategy that leverages real-world data to evaluate the performance of an SER system. The proposed strategy employs data mining algorithms to identify key performance patterns of the SER system, quantify its relationship with external variables, and assess the impact of the recovered heat on the heat supply from the main DH substation. The novelties of this study include 1) development of a two-step clustering analysis using multi-dimensional clustering and temporal clustering techniques to reveal system behaviors and their impact on building energy usage, allowing for a deep understanding of system performance and potential areas for improvement; 2) development of a new performance metric that considers the interactions among different indicators to provide a holistic approach to evaluating the performance of the SER system; and 3) quantification of external factors on various performance patterns of SER systems using a classification and regression trees (CART) model to discover the reasons for performance variation and identify energy saving opportunities for further performance improvement. The study provides a strategy that combines unsupervised data mining (clustering analysis) and supervised data mining (CART analysis) to reveal the operational characteristics and performance of the SER system. The findings of the study can serve as a valuable reference for the development of integrated energy systems to enable efficient integration of waste heat or renewable energy for increased energy flexibility and reduced carbon emissions.

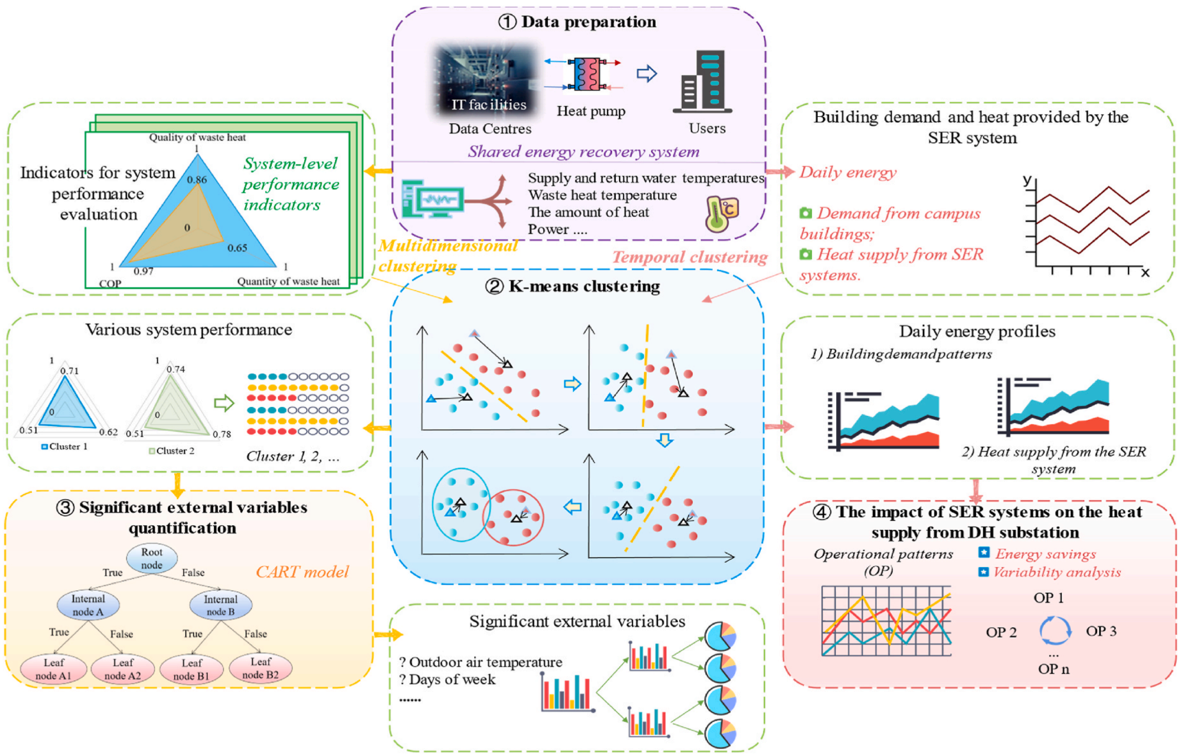


Fig. 2. Outline of the data mining strategy for evaluating the performance of SER systems.

2. A new data mining strategy for performance assessment of shared energy recovery systems

2.1. Outline of the strategy

A new data mining strategy for evaluating the performance of SER systems is presented in Fig. 2. This strategy consists of four steps. In the first step, the measured time-series data from the SER system and the corresponding weather data are collected. The data needs to be first processed to remove outliers and deal with missing data [22]. In this study, the three-sigma rules were used to detect and remove outliers, and linear interpolation was used to fill a small proportion of the missing data [23]. In the second step, a K-means algorithm is employed to perform multidimensional clustering and temporal clustering. Multidimensional clustering is applied to the three performance indicators to identify the typical performance of the SER system. Temporal clustering is applied to daily energy profiles, such as thermal energy demand from buildings and thermal energy provided by the SER system, to identify representative thermal energy use and thermal energy generation patterns. In the third step, based on the multidimensional clustering results, the CART model is used to generate a decision tree to further quantify the influential external variables on the system performance and investigate the reasons for performance variation. Lastly, the influence of the SER system on the thermal energy supply from other heat sources, such as the main DH substation, is studied based on temporal clustering results. This includes the identification of operational patterns, determination of energy savings of the main DH system using recovered heat, and assessment of the variability of daily energy supply from the main DH substation.

The essential steps included in this strategy are detailed in the following sections. Sections 2.2 and 2.3 explain the process for identifying significant indicators of SER system performance, along with the application of the K-means algorithm for multi-dimensional and temporal clustering. Section 2.4 discusses the association of system performance clusters with external variables by employing a CART model. Lastly, Section 2.5 identifies operational patterns and evaluates the impact of SER systems on the operation of the DH substation.

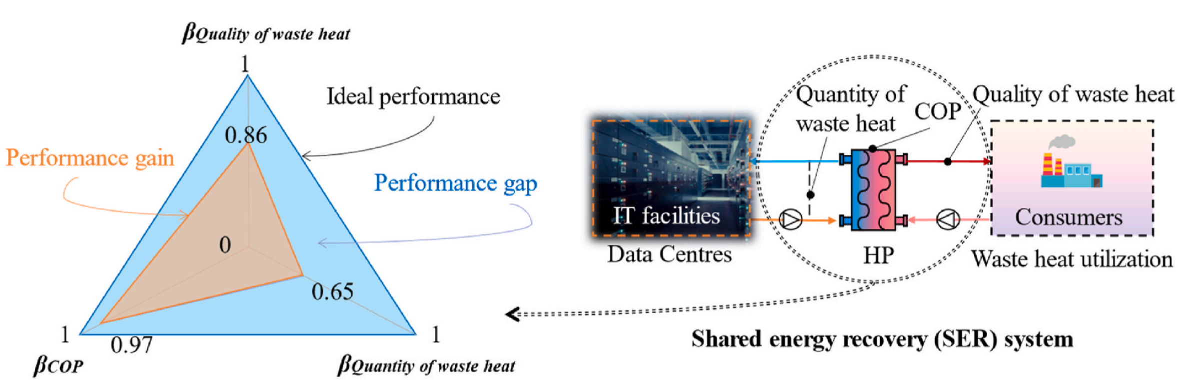


Fig. 3. Illustration of three indicators proposed for performance evaluation of SER systems integrated with data centres.

2.2. Indicators for performance assessment of SER systems

Based on the collected data and domain knowledge, three key indicators of the quality of waste heat supplied by the SER system, the quantity of waste heat shared from the data centre, and the COP of the heat pumps are used to characterize the system performance. The quality of waste heat plays a crucial role in determining the most effective solution to reuse waste heat recovered from data centres to improve energy efficiency. In this particular study, the temperature level of the waste heat was chosen as the quality indicator. The quantity of waste heat (i.e. the instantaneous waste heat recovery rate) from the data centre can represent the amount of waste heat that can be collected from the data centre, which is measured by the heat flow rate, i.e. the amount of waste heat generated per unit of time in this study. Furthermore, COP is used to indicate the efficiency of heat pumps and it represents the ratio of useful heat output to the electricity use of the heat pumps. It is noted that the electricity use of the circulation pumps was not considered in the COP calculation as constant-speed water pumps were used in this study.

Fig. 3 illustrates the interconnection of these three indicators in an SER system, in which heat pumps play an essential role in sharing thermal energy, wherein one side is connected to the cooling system of the data centre and the other side is connected to the consumers (e.g. DH networks) to reuse the waste heat collected. The blue triangle on the left side is the ideal performance of the SER system, where the maximum value of each indicator is one. The orange triangle represents the real performance of the system. The data for each vertex is the ratio of the average value to the maximum value of each indicator, as described in Eq. (1).

$$\beta_i = \frac{\alpha_i}{\alpha_{max}} \tag{1}$$

where α_i indicates the average value of each indicator in that cluster, α_{max} denotes the maximum value of each indicator in the same cluster, and β_i presents the ratio of the average value to the maximum value of each indicator.

2.3. K-means clustering

In this study, the K-means algorithm [24] is utilized to group motifs of the SER performance based on the indicators used, and to cluster daily energy profiles of building demand and waste heat collected from the SER system.

K-means clustering is an iterative process that can minimize the intra-cluster inertia criterion as defined by Eq. (2).

$$C(P,\mu) = \sum\nolimits_{i=1}^{n} \sum_{X_i \in P_k} \|X_i\text{-}\mu_k\|^2 \tag{2}$$

where P=(P$_1$, P$_2$, …, P$_k$) is the set of clusters, $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ is the set of cluster centres, and $||\cdot||$ is the L_2 norm associated with the distance metric.

After randomly selecting initial centroids, the Euclidean distance between each data point and the nearest centroid is calculated, by assigning each point to its closest cluster centre. The centroids are then updated with new values. This iterative process continues until the centroids are not changed anymore [25]. The motifs of the SER system performance (three indicators) were identified through multidimensional clustering, while the daily energy profiles (building thermal energy demand and heat supply from the SER system) were determined through temporal clustering. They both are briefly presented in the following sections.

2.3.1. Multidimensional clustering for identifying performance motifs of the SER system

Multidimensional clustering is a technique used to group data points based on the similarity across multiple dimensions or features [26,27]. In this method, each data point is represented as a vector, where each dimension corresponds to a specific feature or attribute. By analyzing the similarity of data points across all dimensions, multidimensional clustering can identify clusters of data points that share similar characteristics. In this study, this clustering method was utilized to distinguish various performance patterns of the SER system.

2.3.2. Temporal clustering for identifying daily energy profiles of the SER system and building thermal energy demand

Temporal clustering is a form of clustering analysis that concentrates on data with a time-related aspect to group similar data points based on their temporal characteristics or attributes [28]. This analysis can identify patterns and relationships in the data over time, which can then be used for prediction and forecast generation, and to develop a deep understanding of the underlying temporal structures in the data [29].

In the context of clustering analysis, the choice of evaluation indices is crucial for determining the optimal number of clusters. In this study, three evaluation indices namely the Silhouette coefficient index, Calinski-Harabasz, and Davies-Bouldin [30], were used to select the optimal number of clusters in a K-means clustering analysis for the identification of representative building energy use patterns and energy provision patterns of the SER system. These indices assessed the distances between data points within clusters (intra-cluster distances) and the distances between data points belonging to different clusters (inter-cluster distances). Low intra-cluster distances and high inter-cluster distances indicated the presence of distinct clusters. The optimal number was selected based on the combination of the clustering analysis and domain expertise, ensuring that the chosen number was appropriate for the specific context.

2.4. Quantification of external variables for the SER system performance

Except for the three indicators used to evaluate the system performance, the results from the previous studies [31,32] showed that external variables such as weather temperature and days of the week (i.e. Monday to Sunday) greatly impact the performance of DH networks. The supply water temperature from the main substation of the DH system to the buildings also impacts the overall performance of the SER system. Therefore, the influence of these three external variables on the SER system was analyzed to capture the SER system characteristics and discover system performance variations by using a CART model [33]. The CART model is well-suited for dealing with continuous features and allows for quantifying the influence of each variable on system performance [34]. This model can classify the target variable into multiple groups according to the explanatory variables by employing a recursive binary data splitting method [35]. In this study, the outdoor temperatures, the days of the week, and the supply water temperature were considered as the explanatory variables, while SER system performance defined by the three selected indicators was used as the target variables for the CART model training. The model was visually represented to illustrate the relationship between external variables and the performance of the SER system.

2.5. Impact of the SER system on the heat supply from the main DH substation

When the SER system is integrated with DH systems, the amount of waste heat collected, and the quality of the waste heat will impact the operation of the DH systems. The potential impact of the SER system on the operational performance of the heat supply from DH systems is therefore analyzed and quantified in this study based on the results from the temporal clustering analysis. Two indicators are introduced to evaluate the above-mentioned impact, namely the daily energy reduction ratio and peak-to-average ratio (PAR) of heat supply, in which the energy reduction ratio measures energy saving potential, while the PAR

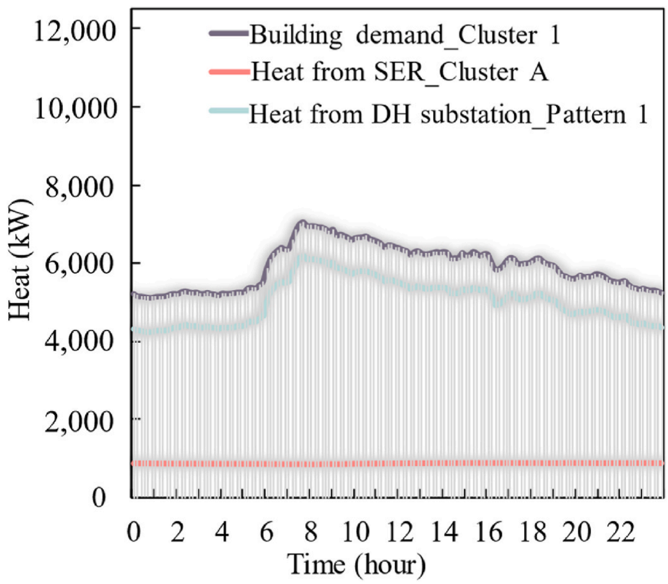


Fig. 4. Example of one operational pattern of building demand and heat supply from the SER system and the main DH substation.

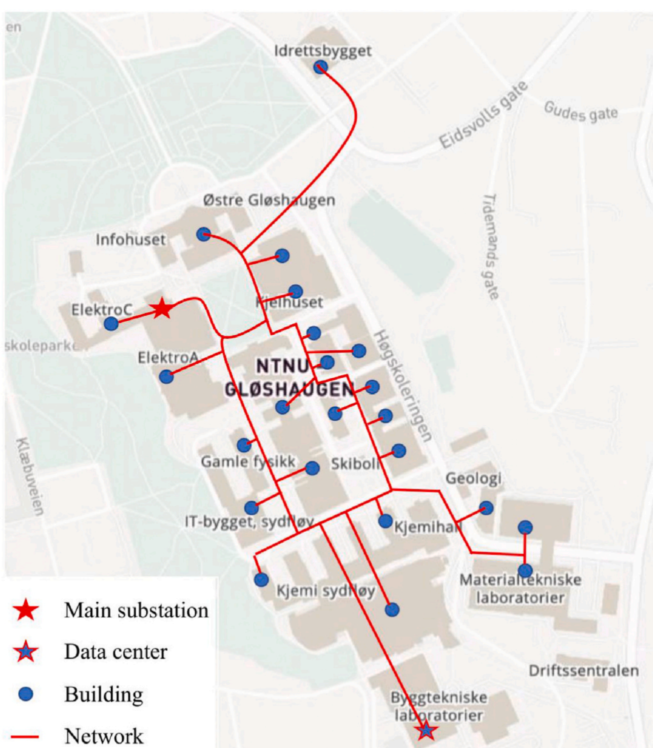


Fig. 5. Topology of the campus DH network integrated with waste heat from a data centre [13].

evaluates the variability of the heat supplied by the main DH substation to the campus buildings.

The temporal clustering can result in x clusters of building demand patterns, and y clusters of waste heat patterns from the SER system. This formulated a total number of x × y possible operational patterns for integrating the SER system with DH networks throughout the period of investigation. It is worth noting that building demand in this study refers to building thermal energy demand. Using Fig. 4 as an example, the black line represented a typical cluster of building demand, the red line

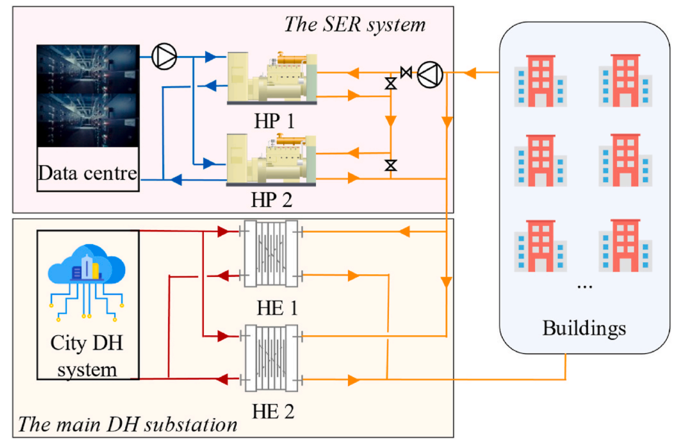


Fig. 6. Configuration of the SER system integrated with a data centre and district heating, where HP indicates heat pump, and HE indicates heat exchanger.

showed a representative cluster of heat provided by the SER system, and the blue line depicted the thermal energy supplied by the main DH substation. As shown in Fig. 4, the building demand increased from 6 a.m. to 8 a.m. and peaked at approximately 7000 kW at around 8 a.m., and then decreased gradually. The variation in building demand appeared to closely align with the heat supply from the main DH substation, while the thermal energy supplied by the SER system was very stable and slightly decreased between 6 and 10 a.m.

3. Results and discussion

3.1. Description of the case study

The performance of the proposed strategy was tested and evaluated using a case study located on a university campus in Norway. The total building area studied is 300,000 m^2, and the main functions of these buildings are teaching buildings, offices, laboratories, and sports centres [36]. Fig. 5 shows the topology of the campus DH networks integrated with a data centre. The campus DH system is connected to the city DH system through a main DH substation. The main DH substation supplies thermal energy for heating and hot water in the campus buildings. A fraction of heat is supplied by the low-grade thermal energy harvested from the data centre of the university.

Fig. 6 illustrates the configuration of the SER system integrated with a data centre and DH network. The entire system includes a data centre, two heat pumps with a design capacity of 680 kW each (serial connection), two water pumps, a main DH substation, and a campus DH network. The two heat pump units were connected in series to increase

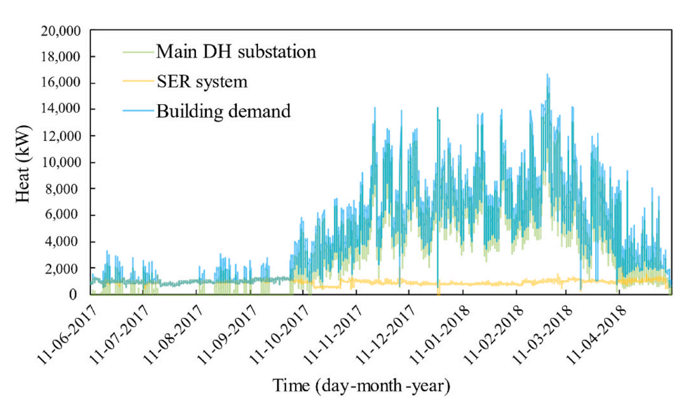


Fig. 7. Building demand, and thermal energy supplied by the SER system and by the main DH substation.

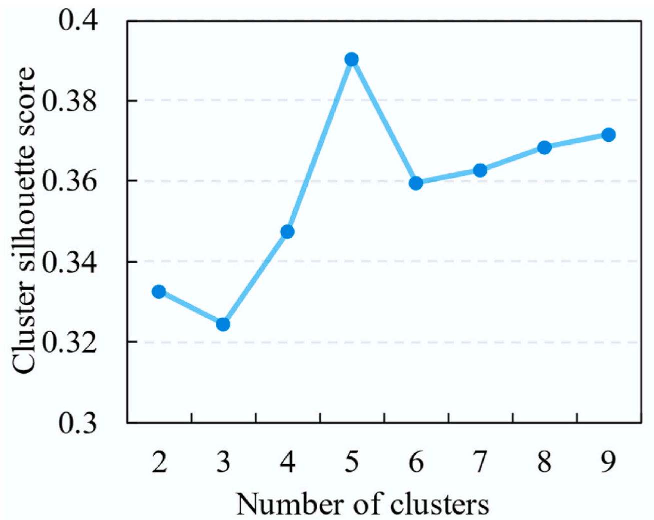


Fig. 8. Determination of the optimum number of clusters.

the temperature of the waste heat recovered from the data centre. The main DH substation consisted of two heat exchangers and two water pumps. Both the SER system and the main DH substation are used to supply thermal energy to buildings. It is noteworthy that the DH network in this study refers to a campus DH system, rather than the whole city heating network.

The data recorded included the supply and return water temperatures and water flow rates of both sides of the heat pumps, the amount of waste heat collected from the data centre and the amount of thermal energy supplied by the SER system to the DH system, the electricity use of the heat pump units, and the outdoor air temperature. The data collected from June 2017 to May 2018 with a time interval of 10 min was used in this analysis.

Fig. 7 presents the recorded data of the amount of thermal energy generated by the SER system, the heat supply from the main DH substation, and the demand of the campus buildings during the data collection period. It is worth noting that the demand for campus buildings encompassed both hot water and heating. The non-heating season was between June and October, while the remaining days were among the heating days. During non-heating months, the thermal energy generated by the SER system can cover the majority of building demand (i.e. hot water), while during the heating months, only around 10 % of building demand was provided by the SER system.

3.2. Performance evaluation of the SER system

3.2.1. Multidimensional clustering for identification of performance motifs of the SER system

The Silhouette coefficient was used to evaluate the quality of the clustering to determine the optimal cluster number and the results are

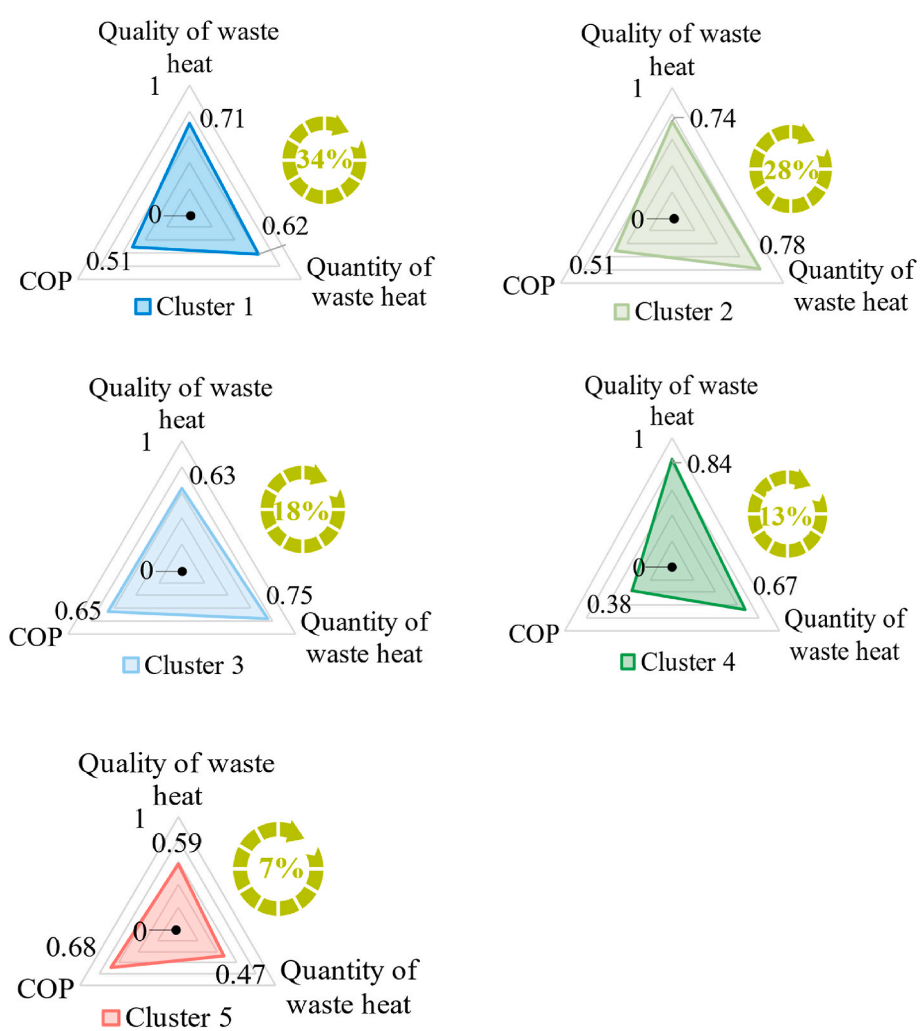


Fig. 9. Typical performance patterns of the SER system and the respective proportion of the operational period in each pattern.

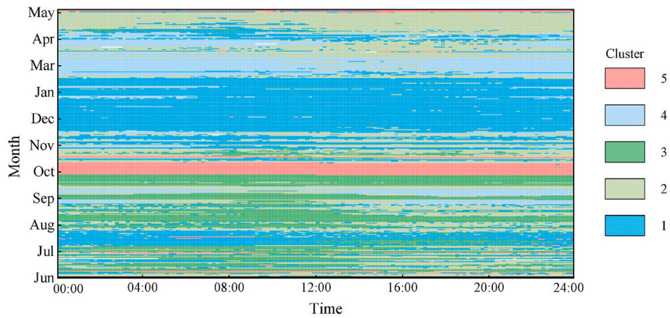


Fig. 10. Heat map of five performance patterns.

shown in Fig. 8. The optimal number of clusters was determined to be five, which was determined based on the largest Silhouette coefficient score. This indicated that the SER system exhibited five distinct performance patterns during this operational period. The performance patterns were then ranked according to the proportion of the operational period in that pattern to the total operational time of concern and were labeled as patterns 1 to 5, from the highest proportion to the lowest one.

The purpose of clustering was to identify the similar quality of waste heat, COP of the heat pump units, and quantity of waste heat as a cluster to represent the typical performance patterns of the system throughout the period. Fig. 9 depicts the five distinct performance patterns of the SER system throughout the operational period. Each cluster was represented by color-coded triangles, which indicated the distinct performance of the system. The green percentages indicated the proportion of time in each pattern during the entire operational period. The data for each vertex of the triangle was expressed as the average value relative to the maximum value in each performance indicator. It was noted that the maximum values observed for the quality of waste heat, heat pump COP, and the quantity of waste heat during the operational period were 80.5 °C, 4.0, and 924 kW, respectively, which corresponded to the ideal maximum value of 1 in Fig. 9.

By comparing these five clusters, it can be seen that the top three clusters accounted for nearly 80 % of the total operational time, representing the typical operational performance. In these top three clusters, the average quality of the waste heat of the system was 55.7 °C, and the quantity of waste heat supplied by the data centre ranged from 573 kW to 720 kW. Although the average temperature might be lower than that of the consumer need, this waste heat was still useful as a stable and reliable heat source for pre-heating the DH network or for other lower temperature heating purposes such as hot water supply. The heat pumps achieved an average COP of 2.2, which was even lower than an air source heat pump under similar working conditions [37], indicating the opportunity for improving heat pump operations. The last two clusters only accounted for 20 % of the total operational period.

As mentioned before, the three performance indicators interact with each other and it is hard to achieve their maximum values simultaneously for all the performance indicators used in each cluster. For example, cluster 1 showed the most prevalent performance during the operational period of concern, accounting for 34 % of the total time. The average quality and quantity of waste heat and COP of the heat pumps were 57.2 °C (corresponding to 0.71 in Fig. 9), 572.9 kW, and 2.0, respectively. The average values were deemed acceptable as no indicators were excessively low compared to other clusters, indicating that the overall system performance under this cluster was satisfactory.

In contrast, cluster 5 was the least frequently occurring pattern during the operational period of interest, representing only 7 % of the total operational time. The COP of the heat pumps in this cluster was 2.7, but the amount of waste heat generated was relatively low, i.e. approximately 50 % of the maximum value. Additionally, the temperature of the waste heat supply from the SER system was only 47.5 °C. This is because only one heat pump was in operation during this period. Therefore, the quality of waste heat could not be raised to a higher level, and there was limited waste heat available to be shared with the end users, although the heat pump showed a higher COP.

Combining cluster 2 and cluster 3 with the characteristics observed in cluster 4, it can be concluded that there was a notable inverse relationship between COP and the quality of waste heat. This exhibited the requirement for system optimization to balance different performance indicators. Fig. 10 is the heat map showing the distribution of five performance patterns during the operational period. Although the performance patterns remained relatively stable without significant alternation between different clusters in a 24-h time frame, periodic occurrences of each performance pattern were observed in different months. The distribution of these different performance patterns appeared to be related to the non-heating season and the heating season. For example, cluster 3 more commonly occurred during the non-heating season. This was because the characteristics of cluster 3 aligned with the conditions of the non-heating season by providing hot water to the end users. Cluster 1 was primarily observed during the heating season when the waste heat collected from the SER system was supplied to the DH system and also used for hot water purposes. The overall performance of cluster 1 was lower than that of cluster 3. This highlighted the potential impact of the DH system on the SER system due to a higher return water temperature to the heat pumps caused by the return-to-return

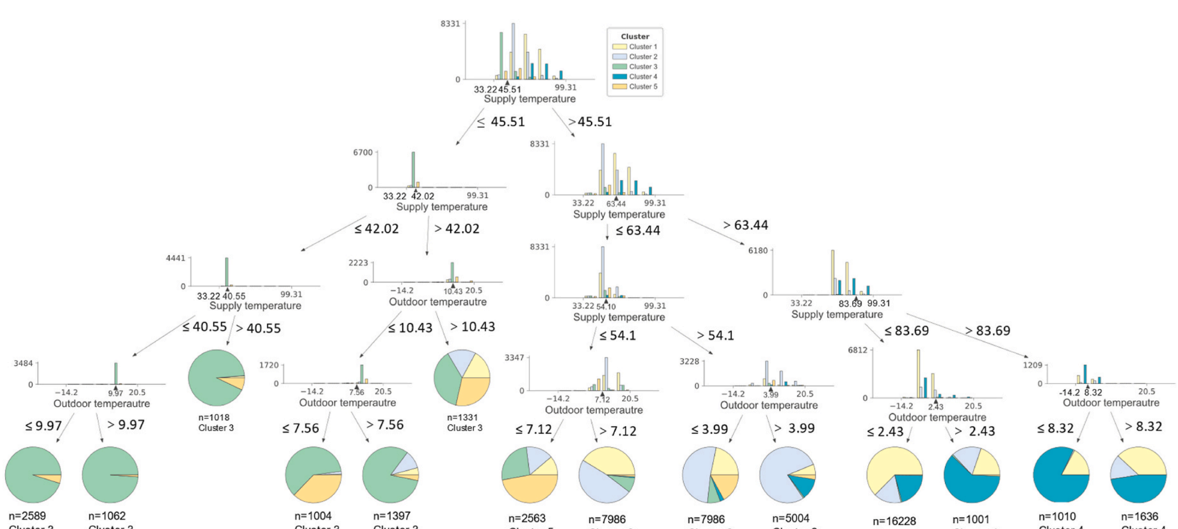
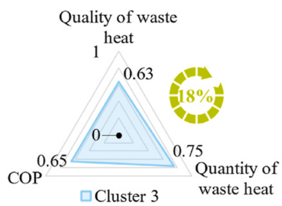
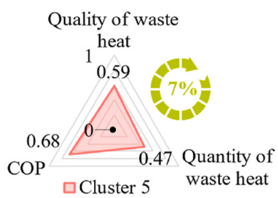
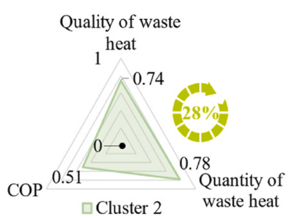
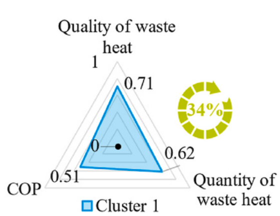
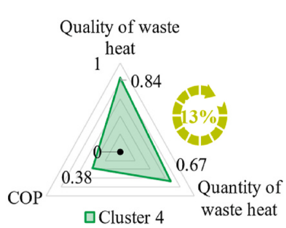


Fig. 11. Impact of external variables on the performance indicators used.

Table 1
Quantitative analysis of two external variables under different system performances.

Supply water temperature from the main DH substation to buildings	Outdoor air temperature	Performance difference
≤45.5		
45.5–54.1	≤7.1	
45.5–54.1 54.1–63.4	>7.1	
63.4–83.7	≤2.4	
63.4–83.7 >83.7	>2.4	

connection of the waste heat [38] and the higher supply temperature of the DH system in the heating season. Moreover, cluster 5 had the lowest occurrence during the operational period and was consistently observed in October, which possibly resulted from equipment failure or human intervention.

3.2.2. *Quantitative analysis of external variables in terms of different performance patterns of the SER system*

As demonstrated in Fig. 9, the SER system exhibits various system performance patterns. Therefore, it is necessary to identify the underlying causes of these differences and the variables that have an impact on the SER system performance. Quantitative analysis can help investigate how external parameters impact different system performance patterns. This information can further be used to optimize the system design and operation to achieve optimal performance under different operating conditions.

In this study, three external variables, outdoor air temperature, supply water temperature from the main DH substation to the buildings, and days of the week, were considered to quantify their impact on the performance of the SER system and the results are shown in Fig. 11. It was noted that in this figure, the term “outdoor temperature” referred to outdoor air temperature (^{o}C) and “supply temperature” represented the temperature (^{o}C) of the supply water from the main DH substation to campus buildings.

It can be seen that among the external variables, the temperature of supply water from the main DH substation to the buildings significantly affected the SER system performance. Additionally, the outdoor air temperature played a critical role in shaping the overall system performance. In contrast, the days of the week showed a relatively minor impact on the various system performance indicators. Hence, the days of the week were not considered in the following analysis.

Table 1 summarizes the results of the quantitative analysis to demonstrate how the external variables, including the temperature of supply water from the main DH substation to buildings, and outdoor air temperature, impacted system performance patterns. When the supply water temperature from the main DH substation to the buildings was below 45.5 °C, the SER system mainly operated in the performance pattern of cluster 3, with an average waste heat quality of 50.0 °C. Combining this with the aforementioned heat map analysis, it can be concluded that during the non-heating season, the SER system was used as a major heat source to supply hot water to the buildings.

When the supply temperature was in the range of 45.5–54.1 °C and the outdoor air temperature was below 7.1 °C, the SER system operated in performance pattern 5. After checking the original data and the performance characteristics depicted in Fig. 10, it was evident that only one heat pump was in operation throughout this specific pattern. Notably, this pattern lasted for approximately two weeks, making it different from the other operational periods. Consequently, the underlying cause of this distinct phenomenon warrants further investigation.

When the outdoor air temperature was above 7.1 °C and the supply temperature ranged from 45.5 to 54.1 °C, or the supply temperature exceeded 54.1 °C but was lower than 63.4 °C, the performance of the SER system was under performance pattern 2. One possible explanation for this was that the end-users required less heating energy to stay warm due to the relatively high outdoor air temperature or supply water temperature.

When the supply temperature ranged from 63.4 °C to 83.7 °C, the SER system was operated under performance pattern 1 if the outdoor air temperature was below 2.4 °C, and it was operated under performance pattern 4 if the outdoor temperature was above 2.4 °C. The system was also operated under performance pattern 4 when the supply temperature exceeded 83.7 °C. This occurrence was likely attributable to an excess of thermal energy supplied to the buildings, potentially leading to overheating issues. To address this issue, some actions could be considered, such as incorporating heating demand flexibility or implementing predictive load adjustment to accurately provide thermal energy in real time and achieve energy savings while maintaining indoor comfort levels.

In summary, the SER system performance is strongly affected by the supply water temperature from the main DH substation to buildings and the outdoor air temperature. A higher water temperature from the main substation can increase the quality of the waste heat, while it can

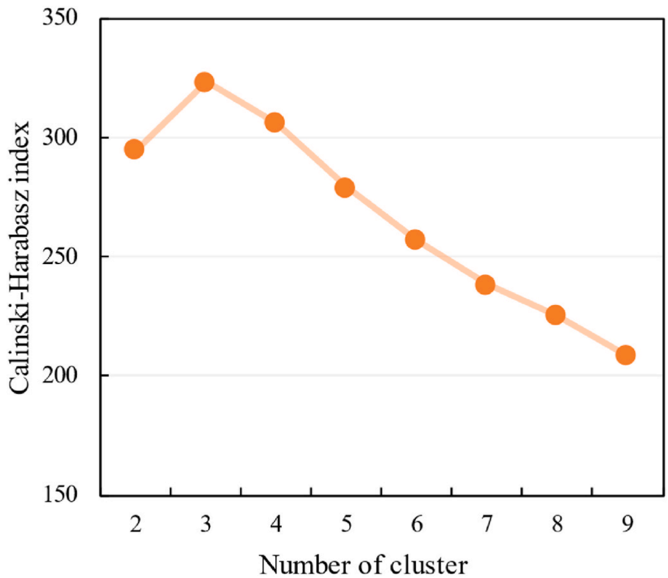


Fig. 12. Optimal number of clusters for building demand.

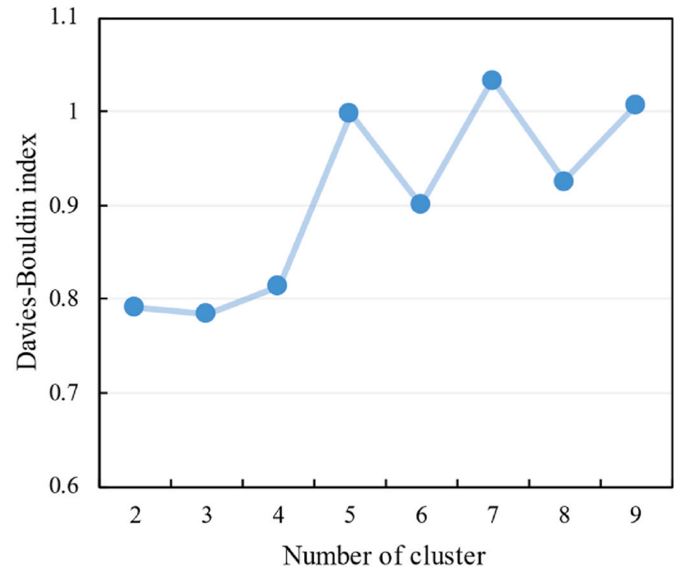


Fig. 14. Optimal number of clusters for thermal energy supplied by the SER system.

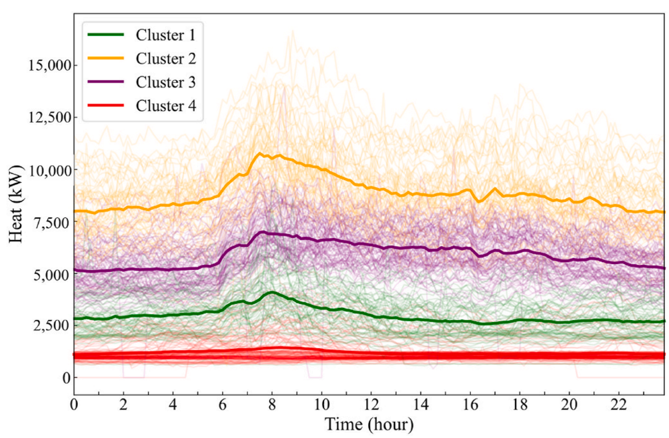


Fig. 13. Four clusters of the building demand.

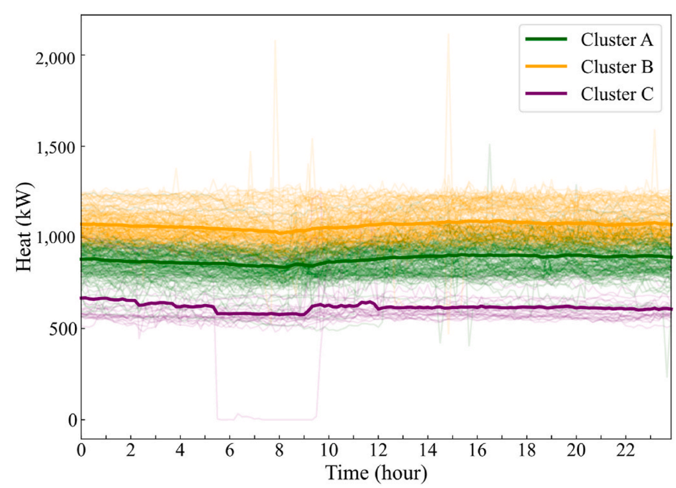


Fig. 15. Three clusters of thermal energy supplied from the SER system.

decrease the COP of the system. To enhance the energy efficiency of the heat pumps and optimize the quality of waste heat produced, it is important to optimize the supply water temperature from the main substations.

Based on the results, it was found that this strategy effectively enabled the generation and comparison of various performance patterns of SER systems and elucidated the interrelationships among different performance indicators. This valuable insight can empower operators to understand and enhance the operational performance of the SER system.

3.3. Identification of operational patterns and assessment of the impact of the SER system on the heat supply from the main DH substation

3.3.1. Temporal clustering analysis

Due to the substantial differences in building demand between the heating and non-heating seasons, the demand patterns during the heating season and those during the non-heating season were respectively characterized. The building demand during the non-heating season was categorized into one group as it was very stable as shown in Fig. 7, while that during the heating season was clustered using the K-means method and the Calinski-Harabasz index to determine the optimal cluster number (Fig. 12). It can be seen that the optimal cluster number was three, as evidenced by the highest value of the Calinski-Harabasz index. Therefore, there were a total of four clusters throughout the entire operational period of concern, including the one during the non-heating season.

As illustrated in Fig. 13, distinct patterns can be observed. Clusters 1, 2, and 3 were characterized by fluctuations with the highest load occurring at around 8 a.m., whereas cluster 4 was comparatively stable. These first three clusters corresponded to the demand during the heating season, while the red pattern (cluster 4) corresponded to the demand during the non-heating season.

The Davies-Bouldin index was used to identify the optimal cluster number of thermal energy provided by the SER system, which was determined based on trial and error tests of different indexes. The optimal number of clusters for the thermal energy provided by the SER system was also three based on the lowest value of the Davies-Bouldin index, as illustrated in Fig. 14.

Fig. 15 illustrates the various clusters of thermal energy provided by the SER system. The thermal energy output of the SER system was found to be consistent and stable throughout the whole period of concern. Clusters A and B exhibited similar and stable trends, while cluster C showed slightly more fluctuations. Specifically, cluster A exhibited the highest thermal energy output, reaching approximately 1100 kW.

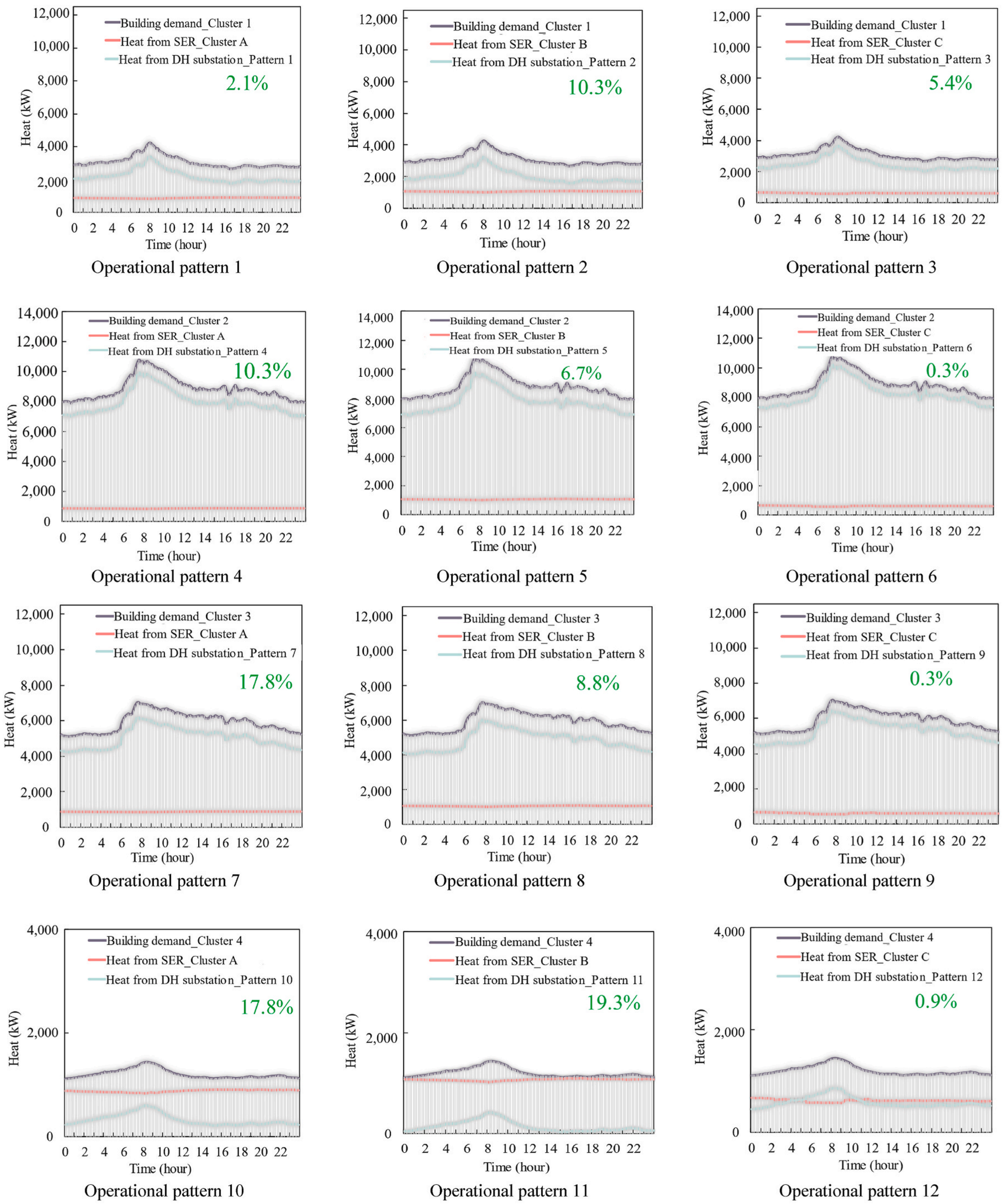


Fig. 16. A total of 12 operational patterns during the operating period.

Meanwhile, cluster C demonstrated the lowest thermal energy generation, ranging from 600 kW to 700 kW. Furthermore, there was a decrease in thermal energy output from the system between 6 a.m. and 9 a.m. This reduction in thermal energy was likely to be caused by the external heat source, e.g. the main DH substation.

3.3.2. Operational pattern identification of the SER system integrated with data centres and DH networks

Based on the above results, the clustering analysis of building demand and thermal energy provided by the SER system resulted in 4 clusters of building demand and 3 clusters of thermal energy. This results in the formation of 12 operational patterns, as shown in Fig. 16.

As demonstrated in Fig. 16, each operational pattern showed unique operational characteristics. Operational patterns 1 to 9 corresponded to the heating season, while operational patterns 10 to 12 represented the non-heating season. During the heating season, the main DH substation served as the primary source of thermal energy for buildings. In contrast, during the non-heating season, the data centre and heat pumps, collectively referred to as the SER system, provided the majority of thermal energy for building operations. Operational patterns 1, 6, 9, and 12 accounted for less than 5 % of the total operational time and can be ignored. All operational patterns involving two heat sources exhibited a mutual impact on the heat output of each other. Specifically, the heat output from the main DH substation tended to increase when that from

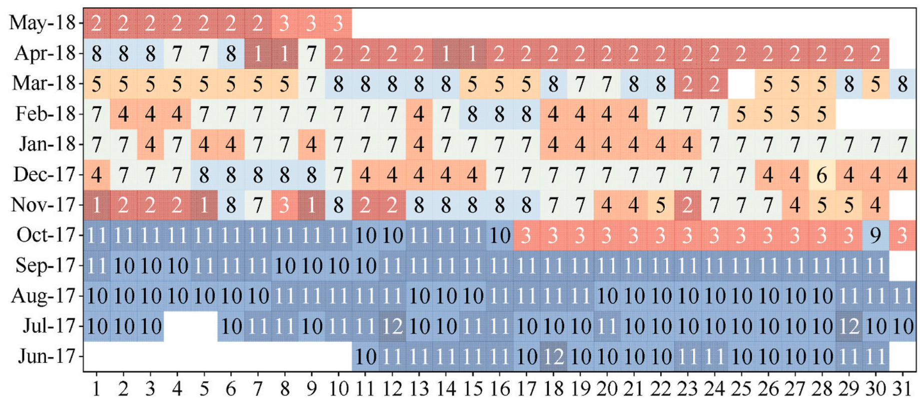


Fig. 17. Heat map of different operational patterns throughout the period.

Table 2
Energy use reduction and variability assessment.

Operational patterns	Average hourly energy use of the main DH substation (kWh)			Peak average ratio of energy use of the main DH substation		
	Scenario 1	Scenario 2	Energy reduction	Scenario 1	Scenario 2	Difference
OP 1	3075.0	2196.3	28.6 % (↓)	1.38	1.55	12.33 % (↑)
OP 2	3075.0	2009.9	34.6 % (↓)	1.38	1.60	16.11 % (↑)
OP 3	3075.0	2458.4	20.1 % (↓)	1.38	1.49	8.05 % (↑)
OP 4	8844.2	7965.5	9.9 % (↓)	1.22	1.25	2.33 % (↑)
OP 5	8844.2	7779.1	12.0 % (↓)	1.22	1.25	2.79 % (↑)
OP 6	8844.2	8227.7	7.0 % (↓)	1.22	1.24	1.73 % (↑)
OP 7	5908.0	5029.2	14.9 % (↓)	1.19	1.23	3.38 % (↑)
OP 8	5908.0	4842.8	18.0 % (↓)	1.19	1.24	4.13 % (↑)
OP 9	5908.0	5291.4	10.4 % (↓)	1.19	1.22	2.41 % (↑)
OP 10	1203.8	325.1	73.0 % (↓)	1.20	1.85	54.88 % (↑)
OP 11	1203.8	138.7	88.5 % (↓)	1.20	2.99	149.56 % (↑)
OP 12	1203.8	587.2	51.2 % (↓)	1.20	1.47	22.68 % (↑)

the SER system tended to decrease.

Fig. 17 shows the distribution of each operational pattern throughout the year. During the non-heating season, operational patterns 10 and 11 alternated in their occurrence. This was due to the primary contribution of the SER system in providing thermal energy, leading to a high energy-saving potential. Operational patterns 1 to 9 occurred during the heating season with operational pattern 7 being the most frequent.

Considering the characteristics of the operational patterns, it was concluded that operational patterns 10 and 11 occurred alternatively. This suggested that when the SER system provided more thermal energy, the main DH substation provided less thermal energy, as the building demand remained stable. Operational pattern 3 occurred on certain days only.

3.3.3. *Impact of the SER system on the heat supply from the main DH substation*

Based on the above results, two scenarios were designed. In Scenario 1, the main DH substation solely covered the building demand. In Scenario 2, the building demand was met by a combination of the SER system and the main DH substation. Two indicators were developed to evaluate heat supply from the main DH substation in the proposed scenarios, namely the energy reduction ratio and PAR of heat supply from the DH substation. The results are presented in Table 2.

Scenario 2 resulted in a significant energy use reduction compared to Scenario 1. During the heating season, the energy-saving potential ranged from 7 % to 35 %. In contrast, during the non-heating season, the energy-saving potential was more than 50 %. This indicated that during the heating season, the main DH substation was the primary heat source, while during the non-heating season, the SER system provided the majority of thermal energy for building demand. Additionally, operational patterns 10 and 11 demonstrated higher energy-saving potential, while operational pattern 6 showed the lowest energy-saving potential of only 7 %.

Regarding the PAR, Scenario 1 outperformed Scenario 2, as the heat supply variability from the main DH substation was lower than that of Scenario 2. Additionally, it can be observed that the main DH substation had less energy variability during the heating season compared to the non-heating season, likely due to the more stable load profile of the main DH substation during the heating season, and the ability to rely on the SER system to cover building demand during the non-heating season.

Compared to Scenario 1, the use of the SER system in Scenario 2 resulted in a significant increase (>55 %) in the PAR of heat supply from the heating substation during most non-heating season operations. This highlighted the need to consider the trade-off between energy consumption reduction and variability increase in the daily energy profile. While waste heat recovery from data centres and heat pumps can lead to increased daily energy profile variability for the main substation, the implementation of appropriate thermal storage and demand flexibility measures can be explored to mitigate this impact.

4. Conclusions

This study proposed a new data mining strategy that utilized multidimensional clustering, temporal clustering, and Classification and Regression Tree model to evaluate the performance of an SER system integrated with a data centre and a district heating system at a university campus. The main findings are as follows. The multidimensional clustering effectively discovered both distinct and predominant performance patterns of the SER system, which exhibited variability attributed to external factors such as outdoor temperature and supply water temperature of the district heating substation. The average Coefficient of Performance for the heat pumps employed within the SER system was around 2.2, slightly lower than those utilized for space air conditioning. However, these heat pumps effectively recovered waste thermal energy from data centres and distributed it to district heating users. The temporal clustering analysis showed that building demand exhibited its peak at around 8 a.m., while the heat supply from the SER system remained stable with a minor decrease from 6 a.m. to 10 a.m. These profiles offered insights into demand flexibility management. Waste heat from the SER system reduced district heating energy use by up to 35 % during heating seasons and over 50 % during non-heating seasons, while it increased variability in heat supply from the district heating substation, with an increase of the peak-to-average ratio by over 55 % during non-heating seasons.

This proposed data mining strategy can also be utilized for the performance evaluation of integrated energy systems. However, the data-driven models used in the strategy should be trained using the data from the target systems, along with the identification of significant performance indicators of the target system. In addition, the effectiveness of the data-driven strategy highly depends on the quantity and quality of the data collected.

Credit author statement

Han Du: Methodology, data analytics, and original draft preparation; **Xinlei Zhou**: Methodology and data analytics; **Natasa Nord**: reviewing and editing; **Yale Carden**: Supervision, reviewing and editing; **Zhenjun Ma**: Supervision, Methodology, reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Programme UNE. Global Status Report for buildings and construction: towards a zero-emission, efficient and resilient buildings and construction sector. 2020. 2020.
[2] Sachs JD, Schmidt-Traub G, Mazzucato M, Messner D, Nakicenovic N, Rockström J. Six transformations to achieve the sustainable development Goals. Nat Sustain 2019;2(9):805–14.
[3] Kim M-H, Kim D, Heo J, Lee D-W. Techno-economic analysis of hybrid renewable energy system with solar district heating for net zero energy community. Energy 2019;187:115916.
[4] Lindhe J, Javed S, Johansson D, Bagge H. A review of the current status and development of 5GDHC and characterization of a novel shared energy system. Sci Technol Built Environ 2022;28(5):595–609.
[5] Abdalla A, Mohamed S, Bucking S, Cotton JS. Modeling of thermal energy sharing in integrated energy communities with micro-thermal networks. Energy Build 2021:248.
[6] Wirtz M, Kivilip L, Remmen P, Müller D. 5th Generation District Heating: a novel design approach based on mathematical optimization. Appl Energy 2020:260.
[7] Murphy AR, Fung AS. Techno-economic study of an energy sharing network comprised of a data centre and multi-unit residential buildings for cold climate. Energy Build 2019;186:261–75.
[8] Ebrahimi K, Jones GF, Fleischer AS. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. Renew Sustain Energy Rev 2014;31:622–38.
[9] Huang P, Copertaro B, Zhang X, Shen J, Löfgren I, Rönnelid M, et al. A review of data centers as prosumers in district energy systems: renewable energy integration and waste heat reuse for district heating. Appl Energy 2020;258.
[10] Davies GF, Maidment GG, Tozer RM. Using data centres for combined heating and cooling: an investigation for London. Appl Therm Eng 2016;94:296–304.
[11] Wahlroos M, Pärssinen M, Manner J, Syri S. Utilizing data center waste heat in district heating – impacts on energy efficiency and prospects for low-temperature district heating networks. Energy 2017;140:1228–38.
[12] Köfinger M, Schmidt RR, Basciotti D, Terreros O, Baldvinsson I, Mayrhofer J, et al. Simulation based evaluation of large scale waste heat utilization in urban district heating networks: optimized integration and operation of a seasonal storage. Energy 2018;159:1161–74.
[13] Li H, Hou J, Hong T, Ding Y, Nord N. Energy, economic, and environmental analysis of integration of thermal energy storage into district heating systems using waste heat from data centres. Energy 2021:219.
[14] Wahlroos M, Pärssinen M, Rinne S, Syri S, Manner J. Future views on waste heat utilization – case of data centers in Northern Europe. Renew Sustain Energy Rev 2018;82:1749–64.
[15] Pan Q, Peng J, Wang R. Experimental study of an adsorption chiller for extra low temperature waste heat utilization. Appl Therm Eng 2019;163.
[16] Araya S, Wemhoff AP, Jones GF, Fleischer AS. Study of a lab-scale organic rankine cycle for the ultra-low-temperature waste heat recovery associated with data centers. J Electron Packag 2021;143(2).
[17] Calikus E, Nowaczyk S, Sant'Anna A, Gadd H, Werner S. A data-driven approach for discovering heat load patterns in district heating. Appl Energy 2019;252: 113409.
[18] Zhang C, Xue X, Zhao Y, Zhang X, Li T. An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems. Appl Energy 2019;253:113492.
[19] Huang K, Yuan J, Zhou Z, Zheng X. Analysis and evaluation of heat source data of large-scale heating system based on descriptive data mining techniques. Energy 2022;251:123834.
[20] Jani HK, Kachhwaha SS, Nagababu G, Das A. Temporal and spatial simultaneity assessment of wind-solar energy resources in India by statistical analysis and machine learning clustering approach. Energy 2022;248:123586.
[21] Zhou X, Lin W, Cui P, Ma Z, Huang T. An unsupervised data mining strategy for performance evaluation of ground source heat pump systems. Sustain Energy Technol Assessments 2021;46.
[22] Zhang S, Zhang L, Zhang X. Clustering based on dynamic time warping to extract typical daily patterns from long-term operation data of a ground source heat pump system. Energy 2022;249:123767.
[23] Khan ZA, Adil M, Javaid N, Saqib MN, Shafiq M, Choi J-G. Electricity theft detection using supervised learning techniques on smart meter data. Sustainability 2020;12(19):8023.
[24] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Convers Manag 2018;165:840–50.
[25] Lumbreras M, Diarce G, Martin K, Garay-Martinez R, Arregi B. Unsupervised recognition and prediction of daily patterns in heating loads in buildings. J Build Eng 2023;65:105732.
[26] Boone-Heinonen J, Gordon-Larsen P, Adair LS. Obesogenic clusters: multidimensional adolescent obesity-related behaviors in the US. Ann Behav Med 2008;36(3):217–30.
[27] Spengler S, Mess F, Mewes N, Mensink GB, Woll A. A cluster-analytic approach towards multidimensional health-related behaviors in adolescents: the MoMo-Study. BMC Publ Health 2012;12:1–8.
[28] Lin W, Orgun MA, Williams GJ. An overview of temporal data mining. AusDM 2002:83–90.
[29] Afzalan M, Jazizadeh F, Eldardiry H. Two-stage clustering of household electricity load shapes for improved temporal pattern representation. IEEE Access 2021;9: 151667–80.
[30] Katarya R, Saini R. Enhancing the wine tasting experience using greedy clustering wine recommender system. Multimed Tool Appl 2022;81(1):807–40.
[31] Wojdyga K. An influence of weather conditions on heat demand in district heating systems. Energy Build 2008;40(11):2009–14.
[32] Chicherin S, Zhuikov A, Kolosov M, Junussova L, Aliyarova M, Yelemanova A. Optimizing building heat consumption: weekday and weekend profiles. Energy Rep 2021;7:166–70.
[33] Zhou H, Zhang J, Zhou Y, Guo X, Ma Y. A feature selection algorithm of decision tree based on feature weight. Expert Syst Appl 2021:164.
[34] Zhou X, Du H, Sun Y, Ren H, Cui P, Ma Z. A new framework integrating reinforcement learning, a rule-based expert system, and decision tree analysis to improve building energy flexibility. J Build Eng 2023;71:106536.
[35] Wang Z, Wang Y, Srinivasan RS. A novel ensemble learning approach to support building energy use prediction. Energy Build 2018;159:109–22.
[36] Guan J, Nord N, Chen S. Energy planning of university campus building complex: energy usage and coincidental analysis of individual buildings with a case study. Energy Build 2016;124:99–111.
[37] Song M, Mao N, Xu Y, Deng S. Challenges in, and the development of, building energy saving techniques, illustrated with the example of an air source heat pump. Therm Sci Eng Prog 2019;10:337–56.
[38] Nord N, Shakerin M, Tereshchenko T, Verda V, Borchiellini R. Data informed physical models for district heating grids with distributed heat sources to understand thermal and hydraulic aspects. Energy 2021;222:119965.